BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e138801
doi: 10.3897/biss.8.138801

OPEN ACCESS

Conference Abstract

Non-Copyrightability of Data in Scientific Publications: A Free-for-All or a Global Commons Partnership?

Jutta Buschbom[‡,§], Laurence Bénichou[|], Donat Agosti[¶], Willi Egloff[¶], Elisa Hermann[#], Mariko Kageyama[¤], Andreas Kroh[«], Patricia Mergen[»,^]

‡ Statistical Genetics, Ahrensburg, Germany
§ Natural History Museum, London, United Kingdom
| Muséum national d'Histoire naturelle, Paris, France
¶ Plazi, Bern, Switzerland
# Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany
¤ Independent Consultant, Seattle, Washington, United States of America
« Naturhistorisches Museum, Vienna, Austria
» Meise Botanic Garden, Meise, Belgium
^ Royal Museum for Central Africa, Tervuren, Belgium

Corresponding author: Jutta Buschbom (jutta.buschbom@nhm.ac.uk)

Received: 08 Oct 2024 | Published: 08 Oct 2024

Citation: Buschbom J, Bénichou L, Agosti D, Egloff W, Hermann E, Kageyama M, Kroh A, Mergen P (2024) Non-Copyrightability of Data in Scientific Publications: A Free-for-All or a Global Commons Partnership? Biodiversity Information Science and Standards 8: e138801. https://doi.org/10.3897/biss.8.138801

Abstract

Scientific publications provide a wealth of peer-reviewed, high-quality data that have been maintained over time, resulting in data persistence. As data repositories with rich provenance information, publications are indispensable sources for the integration and extension of networks of interlinked Findable, Accessible, Interoperable and Reusable (FAIR*[1]) bio/geodiversity data. In this way, they form pivotal fact- and knowledge-based contributions to applications that address the biodiversity crisis.

The mobilization of data preserved in scientific publications is hindered, however, by distinct copyright legislation contexts for publications versus the data that they contain. Moreover, legislations concerning copyright continue to lack harmonization across jurisdictions, their interpretation is difficult, and the applicable legal national scope can be uncertain.

© Buschbom J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

We clarify and highlight that data within scientific publications are not copyrightable and thus can be openly and freely reused once legal access has been gained to their enclosing publication*[2]. To ensure that publications are as accessible as possible, a joint statement supported by the Biodiversity Heritage Library (BHL), the Consortium of European Taxonomic Facilities (CETAF) and the Society for the Preservation of Natural History Collections (SPNHC) (Benichou et al. 2023) recommends that authors and publishers make their works as accessible as possible by using a CC-BY license or preferably waive copyright (CC0) to their publications. Explicitly associating a public domain mark (PDM, e.g., the PDM from Creative Commons) to their published data, provides users with certainty about reusability.

Yet, by setting works and bio/geodiversity data into the public domain, they do not become a free-for-all. We stress that data need to be associated with clear provenance information in alignment with scientific best practices and the scientific community's social norms. This includes providing detailed attribution to authors of cited works and reused data. Proposed data governance labels, for example, modeled after the Local Contexts labels developed by the international Indigenous Peoples and Local Communities (IPLC) community, would enable authors to communicate social and ethical contexts and applicable rules to data users for ensuring the sustainability of a shared environmental and data commons. Categories of Local Contexts labels that are of interest and applicable in the sciences are, for example, those that communicate (1) correct citation information and ask for attribution when knowledge and/or data are reused (Traditional Knowledge label (TK) Attribution), (2) an interest in being recognized and acknowledged due to a significant relationship with and responsibility for samples and data (Biocultural label (BC) Provenance), (3) the verification of the data and their context following a community protocol (TK Verified), (4) that non-commercial use (TK Non-Commercial/BC Non-Commercial) or (5) outreach activities (TK Outreach/BC Outreach) are generally permitted, while for other uses direct contact and engagement is required, or (6) an openness to collaboration and partnerships (TK Collaboration/BC Collaboration).

There are concerns about the tension between the goal of achieving open data (e.g., Anonymous 2014) to enable and promote open science (e.g., UNESCO 2021) and, at the same time, imposing restrictions on these data in the form of governance labels. Furthermore, while the reference of the publication through which data are published, as well as more specifically bibliographic references cited for specific data within the publication, provide sufficient information for attribution and provenance, much more fine-grained and nuanced contextual information (e.g., in the form of metadata) is needed for assuring responsible reuse. Such context-providing metadata unlock the full potential of the data and enable their reusability. This can be done using machine-actionable markup tags in combination with human-readable labels that inform machines and human users about the semantics of the data as well as their ethical and social dimensions that govern responsible and sustainable reuse.

Future work is needed to discover, differentiate and define the quality and scope of the appropriate contexts that are necessary and sufficient for being able to fully and responsibly reuse the data in different situations.

Keywords

FAIR, linked biodiversity data, public domain mark, scientific best practices, code of conduct, data governance labels

Presenting author

Jutta Buschbom

Presented at

SPNHC-TDWG 2024

Acknowledgements

The authors would like to give special recognition to Constance Rinaldo, Washington D.C., USA (https://orcid.org/0000-0002-8339-728X), a member of the Biodiversity Heritage Library. Connie's engagement moved the beginning process forward and she was contributing to the preparation of the joint-workshop at the TDWG meeting in 2022 before she passed away on October 27, 2022. It was our aim to continue the development of the workshop and its resulting publications in Connie's spirit.

Funding program

This initiative was supported by the BiCIKL project, which receives funding from the European Union's Horizon 2020 Research and Innovation Action under grant agreement No 101007492.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Anonymous (2014) The Bouchout Declaration for Open Biodiversity Knowledge Management. Accessed December 28, 2023. URL: https://www.bouchoutdeclaration.org/
- Benichou L, Agosti D, Egloff W, Hermann E, Kageyama M, Mergen P, Rinaldo C, Buschbom J (2023) Joint statement by CETAF, SPNHC and BHL on DATA within scientific publications: clarification of [non]copyrightability. Research Ideas and Outcomes 9 https://doi.org/10.3897/rio.9.e115466

- UNESCO (2021) UNESCO Recommendation on Open Science. Paris, France. https://doi.org/10.54677/MNMH8546
- Ziaja GM (2024) The text and data mining opt-out in Article 4(3) CDSMD: Adequate veto right for rightholders or a suffocating blanket for European artificial intelligence innovations? Journal of Intellectual Property Law & Practice 19 (5): 453-459. [In English]. https://doi.org/10.1093/jiplp/jpae025

Endnotes

*1 See for example, these resources for more information on FAIR data and their implementation: the FAIR Principles of GO FAIR, and the FAIR Cookbook of the EU Elixir-project.

*2 Recent developments show that Article 4(3) of the EU Directive 2019/790 on Copyright, which introduces an opt-out option for rights-holders, blocking text and data mining, is widely discussed in scientific and business contexts (e.g., Ziaja 2024). We will continue to monitor the developments and will report impacts of the evolving legal landscape on the work of our communities.